EFFICIENT FREQUENCY-DOMAIN IMPLEMENTATION OF SPEECH DISTORTION WEIGHTED MULTI-CHANNEL WIENER FILTERING FOR NOISE REDUCTION

Simon Doclo, Ann Spriet, Marc Moonen

Katholieke Universiteit Leuven, Dept. of Electrical Engineering (ESAT-SCD)
Kasteelpark Arenberg 10, 3001 Leuven, Belgium
{simon.doclo,ann.spriet,marc.moonen}@esat.kuleuven.ac.be

ABSTRACT

A stochastic gradient implementation of a generalised multi-microphone noise reduction scheme, called the Spatially Pre-processed Speech Distortion Weighted Multi-channel Wiener Filter (SP-SDW-MWF), has recently been proposed in [1]. In order to compute a regularisation term in the filter update formulas, data buffers are required in this implementation, resulting in a large memory usage. This paper shows that by approximating this regularisation term in the frequency-domain the memory usage (and the computational complexity) can be reduced drastically. Experimental results demonstrate that this approximation only gives rise to a limited performance difference and that hence the proposed algorithm preserves the robustness benefit of the SP-SDW-MWF over the GSC (with Quadratic Inequality Constraint).

1. INTRODUCTION

Noise reduction algorithms in hearing aids and cochlear implants are crucial for hearing impaired persons to improve speech intelligibility in background noise. Multi-microphone systems exploit spatial in addition to temporal and spectral information of the desired and noise signals and are hence preferred to single-microphone systems. For small-sized arrays such as in hearing instruments, multi-microphone noise reduction however goes together with an increased sensitivity to errors in the assumed signal model such as microphone mismatch, reverberation, etc.

In [2] a generalised noise reduction scheme, called the Spatially Pre-processed Speech Distortion Weighted Multi-channel Wiener Filter (SP-SDW-MWF), has been proposed (cf. Section 2). It encompasses both the Generalised Sidelobe Canceller (GSC) and the MWF [3, 4] as extreme cases and allows for in-between solutions such as the Speech Distortion Regularised GSC (SDR-GSC). By taking speech distortion explicitly into account in the design criterion of the adaptive stage, the SP-SDW-MWF (and the SDR-GSC) add robustness against model errors to the GSC. Compared to the widely studied GSC with Quadratic Inequality Constraint (QIC) [5], the SP-SDW-MWF achieves a better noise reduction performance for a given maximum speech distortion level.

In [1] cheap (time-domain and frequency-domain) stochastic gradient algorithms for implementing the SDW-MWF have been presented. These algorithms however require large data buffers for calculating a regularisation term required in the filter update formulas (cf. Section 3). By approximating this regularisation term in the frequency-domain, (diagonal) speech and noise correlation matrices need to be stored, such that the memory usage is decreased drastically, while also the computational complexity is further reduced. Experimental results using a hearing aid demonstrate that this approximation results in a small - positive or negative - performance difference, such that the proposed algorithm preserves the robustness benefit of the SP-SDW-MWF over the QIC-GSC, while its computational complexity and memory usage are comparable to the NLMS-based algorithm for QIC-GSC.

Simon Doclo is a postdoctoral researcher funded by KULeuven-BOF. This work was supported in part by F.W.O. Project G.0233.01, *Signal processing and automatic patient fitting for advanced auditory prostheses*, I.W.T. Project 020540, *Performance improvement of cochlear implants by innovative speech processing algorithms*, I.W.T. Project 020476, *Sound Management System for Public Address systems (SMS4PA)*, Concerted Research Action GOA-MEFISTO-666, Interuniversity Attraction Pole IUAP P5-22, and was partially sponsored by Cochlear.

2. SPATIALLY PRE-PROCESSED SDW-MWF

The SP-SDW-MWF, depicted in Figure 1, consists of a *fixed spatial pre-processor*, i.e. a fixed beamformer $\mathbf{A}(z)$ and a blocking matrix $\mathbf{B}(z)$, and an *adaptive Speech Distortion Weighted Multi-channel Wiener Filter (SDW-MWF)* [2]. Note that this structure strongly resembles the GSC [5, 6, 7], where the standard adaptive filter has been replaced by an adaptive SDW-MWF.

The desired speaker is assumed to be in front of the microphone array (having M microphones), and an endfire array is used. The fixed beamformer creates a so-called speech reference $y_0[k] = x_0[k] + v_0[k]$ (with $x_0[k]$ and $v_0[k]$ respectively the speech and the noise component of $y_0[k]$) by steering a beam towards the front, whereas the blocking matrix creates $M-1$ so-called noise references $y_i[k] = x_i[k] + v_i[k], i = 1 \dots M-1$, by steering zeroes towards the front. During *speech-periods* these references consist of speech+noise, i.e. $y_i[k] = x_i[k] + v_i[k]$, whereas during *noise-only-periods* the noise components $v_i[k]$ are observed. We assume that the second-order statistics of the noise are sufficiently stationary such that they can be estimated during noise-only-periods and used during subsequent speech-periods. This requires the use of a voice activity detection (VAD) mechanism.

Let N be the number of input channels to the multi-channel Wiener filter in Figure 1 ($N = M$ if $\mathbf{w}_0$ is present, $N = M-1$ otherwise). Let the FIR filters $\mathbf{w}_i[k]$ have length L, and consider the L-dimensional data vectors $\mathbf{y}_i[k]$, the NL-dimensional stacked filter $\mathbf{w}[k]$ and the NL-dimensional stacked data vector $\mathbf{y}[k]$, defined as

$$\begin{aligned}
\mathbf{y}_i[k] &= \begin{bmatrix} y_i[k] & y_i[k-1] & \dots & y_i[k-L+1] \end{bmatrix}^T, && (1)\\
\mathbf{w}[k] &= \begin{bmatrix} \mathbf{w}_{M-N}^T[k] & \mathbf{w}_{M-N+1}^T[k] & \dots & \mathbf{w}_{M-1}^T[k] \end{bmatrix}^T, && (2)\\
\mathbf{y}[k] &= \begin{bmatrix} \mathbf{y}_{M-N}^T[k] & \mathbf{y}_{M-N+1}^T[k] & \dots & \mathbf{y}_{M-1}^T[k] \end{bmatrix}^T, && (3)
\end{aligned}$$

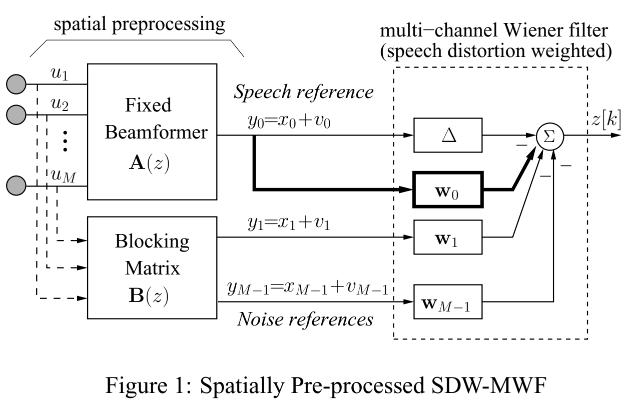

Figure 1: Spatially Pre-processed SDW-MWF

with ^{T} denoting transpose. The data vector $\mathbf{y}[k]$ can be decomposed into a speech component and a noise component, i.e. $\mathbf{y}[k] = \mathbf{x}[k] + \mathbf{v}[k]$, with $\mathbf{x}[k]$ and $\mathbf{v}[k]$ defined similarly as in (3).

The goal of the SDW-MWF is to provide an estimate of the noise component $v_0[k-\Delta]$ in the speech reference by minimising the cost function [2]

$$\boxed{J(\mathbf{w}[k]) = \frac{1}{\mu}\underbrace{\mathscr{E}\left\{\left|\mathbf{w}^T[k]\mathbf{x}[k]\right|^2\right\}}_{\varepsilon_x^2} + \underbrace{\mathscr{E}\left\{\left|v_0[k-\Delta] - \mathbf{w}^T[k]\mathbf{v}[k]\right|^2\right\}}_{\varepsilon_v^2}} \tag{4}$$

where ε_x^2 represents the speech distortion energy, ε_v^2 represents the residual noise energy and the parameter $\mu \in [0,\infty)$ provides a trade-off between noise reduction and speech distortion [3]. As depicted in Figure 1, the noise estimate $\mathbf{w}^T[k]\mathbf{y}[k]$ is then subtracted from the speech reference in order to obtain the enhanced output signal $z[k]$. Depending on the setting of μ and the presence/absence of the filter $\mathbf{w}_0$ on the speech reference, different algorithms are obtained:

- Without $\mathbf{w}_0$, we obtain the Speech Distortion Regularised GSC (SDR-GSC), where the standard ANC design criterion (i.e. minimising the residual noise energy ε_v^2) is supplemented with a *regularisation term* $\frac{1}{\mu}\varepsilon_x^2$ that takes into account speech distortion due to signal model errors. For $\mu = \infty$, the standard GSC is obtained. In [2] it has been shown that in comparison with the QIC-GSC, the SDR-GSC obtains a better noise reduction for small model errors, while guaranteeing robustness against large model errors.
- With $\mathbf{w}_0$, we obtain the SP-SDW-MWF (for $\mu = 1$, we obtain an MWF, where the output signal $z[k]$ is the MMSE estimate of the speech component $x_0[k-\Delta]$ in the speech reference). In [2] it has been shown that in comparison with the SDR-GSC, the performance of the SP-SDW-MWF is even less affected by signal model errors.

Different implementations exist for computing and updating the filter $\mathbf{w}[k]$. In [3, 4] recursive matrix-based implementations (using GSVD and QRD) have been proposed, while in [1] cheap stochastic gradient implementations have been developed.

3. STOCHASTIC GRADIENT ALGORITHM (SG)

3.1 Time-Domain (TD) implementation

In [1] a stochastic gradient algorithm in the time-domain has been developed for minimising the cost function $J(\mathbf{w}[k])$ in (4), i.e.

$$\mathbf{w}[k+1] = \mathbf{w}[k] + \rho\left[\mathbf{v}[k](v_0[k-\Delta] - \mathbf{v}^T[k]\mathbf{w}[k]) - \mathbf{r}[k]\right] \tag{5}$$

$$\mathbf{r}[k] = \frac{1}{\mu}\mathbf{x}[k]\mathbf{x}^T[k]\mathbf{w}[k] \tag{6}$$

$$\rho = \frac{\rho'}{\mathbf{v}^T[k]\mathbf{v}[k] + \frac{1}{\mu}\mathbf{x}^T[k]\mathbf{x}[k] + \delta}\,, \tag{7}$$

with ρ the normalised step size of the adaptive algorithm, δ a small positive constant, and $\mathbf{w}[k]$, $\mathbf{v}[k]$, $\mathbf{x}[k]$ and $\mathbf{r}[k]$ NL-dimensional vectors. For $1/\mu = 0$ and no filter $\mathbf{w}_0$ present, (5) reduces to an NLMS-type update formula often used in GSC, *operated during noise-only-periods* [6, 7]. For $1/\mu \neq 0$, the additional regularisation term $\mathbf{r}[k]$ limits speech distortion due to signal model errors.

In order to compute (6), knowledge about the (instantaneous) correlation matrix $\mathbf{x}[k]\mathbf{x}^T[k]$ of the clean speech signal is required, which is obviously not available. In order to avoid the need for calibration, it is suggested in [1] to store L-dimensional speech+noise-vectors $\mathbf{y}_i[k]$, $i = M-N\ldots M-1$ during speech-periods in a circular *speech+noise-buffer* $\mathbf{B}_y \in \mathbb{R}^{NL\times L_y}$ (similar as in [8])[1] and to adapt the filter $\mathbf{w}[k]$ using (5) during *noise-only-periods*, based on approximating the regularisation term in (6) by

$$\mathbf{r}[k] = \frac{1}{\mu}\left[\mathbf{y}_{B_y}[k]\mathbf{y}_{B_y}^T[k] - \mathbf{v}[k]\mathbf{v}^T[k]\right]\mathbf{w}[k]\,, \tag{8}$$

with $\mathbf{y}_{B_y}[k]$ a vector from the circular speech+noise-buffer $\mathbf{B}_y$. However, this estimate of $\mathbf{r}[k]$ is quite bad, resulting in a large excess error, especially for small μ and large ρ'. Hence, it has been suggested to use an estimate of the average clean speech correlation matrix $\mathscr{E}\{\mathbf{x}[k]\mathbf{x}^T[k]\}$ in (6), such that $\mathbf{r}[k]$ can be computed as

$$\mathbf{r}[k] = \frac{1}{\mu}(1-\bar{\lambda})\sum_{l=0}^{k}\bar{\lambda}^{k-l}\left[\mathbf{y}_{B_y}[l]\mathbf{y}_{B_y}^T[l] - \mathbf{v}[l]\mathbf{v}^T[l]\right]\cdot\mathbf{w}[k]\,, \tag{9}$$

with $\bar{\lambda}$ an exponential weighting factor and the step size ρ in (7) now equal to

$$\rho = \frac{\rho'}{\mathbf{v}^T[k]\mathbf{v}[k] + \frac{1}{\mu}(1-\bar{\lambda})\sum_{l=0}^{k}\bar{\lambda}^{k-l}\left|\mathbf{y}_{B_y}^T[l]\mathbf{y}_{B_y}[l] - \mathbf{v}^T[l]\mathbf{v}[l]\right| + \delta}\,.$$

For *stationary noise* a small $\bar{\lambda}$, i.e. $1/(1-\bar{\lambda}) \sim NL$, suffices. However, in practice the speech and the noise signals are often *spectrally highly non-stationary* (e.g. multi-talker babble noise), whereas their *long-term* spectral and spatial characteristics usually vary more slowly in time. Spectrally highly non-stationary noise can still be spatially suppressed by using an estimate of the *long-term* correlation matrix in $\mathbf{r}[k]$, i.e. $1/(1-\bar{\lambda}) \gg NL$.

In order to avoid expensive matrix operations for computing (9), it is assumed in [1] that $\mathbf{w}[k]$ varies slowly in time, i.e. $\mathbf{w}[k] \approx \mathbf{w}[l]$, such that (9) can be approximated without matrix operations as

$$\mathbf{r}[k] = \bar{\lambda}\mathbf{r}[k-1] + (1-\bar{\lambda})\frac{1}{\mu}\left[\mathbf{y}_{B_y}[k]\mathbf{y}_{B_y}^T[k] - \mathbf{v}[k]\mathbf{v}^T[k]\right]\mathbf{w}[k]\,. \tag{10}$$

However, as will be shown in the next paragraph, this assumption is actually not required in a frequency-domain implementation.

3.2 Efficient Frequency-Domain (FD) implementation

In [1] the SG-TD algorithm has been converted to a frequency-domain implementation by using a block-formulation and overlap-save procedures (similar to standard FD adaptive filtering techniques [9]). However, the SG-FD algorithm in [1] (**Algorithm 1**) requires the storage of large data buffers (with typical buffer lengths $L_y = 10000\ldots20000$). A substantial memory (and computational complexity) reduction can be achieved by the following two steps:

- When using (9) instead of (10) for calculating the regularisation term, *correlation matrices* instead of data buffers need to be stored. The FD implementation of the total algorithm is then summarised in **Algorithm 2**, where $2L\times 2L$-dimensional speech and noise correlation matrices $\mathbf{S}_y^{ij}[k]$ and $\mathbf{S}_v^{ij}[k]$, $i,j = M-N\ldots M-1$ are used for calculating the regularisation term $\mathbf{R}_i[k]$ and (part of) the step size $\mathbf{\Lambda}[k]$. These correlation matrices are updated respectively during speech-periods and noise-only-periods[2]. However, this first step does not necessarily reduce the memory usage (NL_y for data buffers vs. $2(NL)^2$ for correlation matrices) and will even increase the computational complexity, since the correlation matrices are not diagonal.
- The correlation matrices in the frequency-domain can be approximated by diagonal matrices, since $\mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}$ in Algorithm 2 can be well approximated by $\mathbf{I}_{2L}/2$ [10]. Hence, the speech and the noise correlation matrices are updated as

$$\mathbf{S}_y^{ij}[k] = \lambda\mathbf{S}_y^{ij}[k-1] + (1-\lambda)\mathbf{Y}_i^H[k]\mathbf{Y}_j[k]/2\,, \tag{11}$$

$$\mathbf{S}_v^{ij}[k] = \lambda\mathbf{S}_v^{ij}[k-1] + (1-\lambda)\mathbf{V}_i^H[k]\mathbf{V}_j[k]/2\,, \tag{12}$$

[1] In [1] it has been shown that storing noise-only-vectors $\mathbf{v}_i[k]$, $i = M-N\ldots M-1$ during noise-only-periods in a circular *noise-buffer* $\mathbf{B}_v \in \mathbb{R}^{ML\times L_v}$ additionally allows adaptation during speech+noise-periods.

[2] When using correlation matrices, filter adaptation can only take place during noise-only-periods, since during speech-periods the desired signal $\mathbf{d}[k]$ cannot be constructed from the noise-buffer $\mathbf{B}_v$ any more.

Algorithm 2 FD implementation (without approximation)

Initialisation and matrix definitions:

$\mathbf{W}_i[0] = [\ 0 \ \ \cdots \ \ 0\ ]^T,\ i = M-N \dots M-1$

$P_m[0] = \delta_m,\ m = 0 \dots 2L-1$

$\mathbf{F} = 2L \times 2L$-dimensional DFT matrix

$\mathbf{g} = \begin{bmatrix} \mathbf{I}_L & \mathbf{0}_L \\ \mathbf{0}_L & \mathbf{0}_L \end{bmatrix},\quad \mathbf{k} = [\ \mathbf{0}_L \ \ \mathbf{I}_L\ ]$

$\mathbf{0}_L = L \times L$ matrix with zeros, $\mathbf{I}_L = L \times L$ identity matrix

For each new block of L samples (per channel):

$\mathbf{d}[k] = [\ y_0[kL-\Delta] \ \ \cdots \ \ y_0[kL-\Delta+L-1]\ ]^T$

$\mathbf{Y}_i[k] = \mathrm{diag}\left\{\mathbf{F}[\ y_i[kL-L] \ \ \cdots \ \ y_i[kL+L-1]\ ]^T\right\}$

Output signal:

$$\mathbf{e}[k] = \mathbf{d}[k] - \mathbf{k}\mathbf{F}^{-1}\sum_{j=M-N}^{M-1}\mathbf{Y}_j[k]\mathbf{W}_j[k], \quad \mathbf{E}[k] = \mathbf{F}\mathbf{k}^T\mathbf{e}[k]$$

If speech detected:

$$\mathbf{S}_y^{ij}[k] = (1-\lambda)\sum_{l=0}^{k}\lambda^{k-l}\mathbf{Y}_i^H[l]\mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}\mathbf{Y}_j[l]$$

If noise detected: $\mathbf{V}_i[k] = \mathbf{Y}_i[k]$

$$\mathbf{S}_v^{ij}[k] = (1-\lambda)\sum_{l=0}^{k}\lambda^{k-l}\mathbf{V}_i^H[l]\mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}\mathbf{V}_j[l]$$

Update formula (only during noise-only-periods):

$$\mathbf{R}_i[k] = \frac{1}{\mu}\sum_{j=M-N}^{M-1}\left[\mathbf{S}_y^{ij}[k] - \mathbf{S}_v^{ij}[k]\right]\mathbf{W}_j[k]$$

$$\mathbf{W}_i[k+1] = \mathbf{W}_i[k] + \mathbf{F}\mathbf{g}\mathbf{F}^{-1}\mathbf{\Lambda}[k]\left\{\mathbf{V}_i^H[k]\mathbf{E}[k] - \mathbf{R}_i[k]\right\}$$

with

$$\mathbf{\Lambda}[k] = \frac{2\rho'}{L}\,\mathrm{diag}\left\{P_0^{-1}[k], \dots, P_{2L-1}^{-1}[k]\right\}$$

$$P_m[k] = \gamma P_m[k-1] + (1-\gamma)\left(P_{v,m}[k] + P_{x,m}[k]\right)$$

$$P_{v,m}[k] = \sum_{j=M-N}^{M-1}\left|V_{j,m}[k]\right|^2,\ P_{x,m}[k] = \frac{1}{\mu}\left|\sum_{j=M-N}^{M-1}S_{y,m}^{jj}[k] - S_{v,m}^{jj}[k]\right|$$

leading to a significant reduction in memory usage (and computational complexity), cf. Section 4, while having a minimal impact on the performance and the robustness, cf. Section 5. We will refer to this algorithm as **Algorithm 3**. This algorithm is in fact quite similar to [11], which is derived directly from a frequency-domain cost function. Some major differences however exist, e.g. in [11] the regularisation term $\mathbf{R}_i[k]$ is absent, the term $\mathbf{F}\mathbf{g}\mathbf{F}^{-1}$ is also approximated by $\mathbf{I}_{2L}/2$ and the speech and the noise correlation matrices are block-diagonal.

4. MEMORY AND COMPUTATIONAL COMPLEXITY

Table 1 summarises the computational complexity and the memory usage for the FD implementation of the QIC-GSC (computed using the NLMS-based Scaled Projection Algorithm (SPA)[3] [5]) and the SDW-MWF (Algorithm 1 and 3). The computational complexity is expressed as the number of operations (i.e. real multiplications and additions (MAC) per second) in MIPS and the memory usage is expressed in kWords. We assume that one complex multiplication is equivalent to 4 real multiplications and 2 real additions and that a $2L$-point FFT of a real input vector requires $2L\log_2 2L$ real MACs

[3]The complexity of the FD GSC-SPA also represents the complexity when the adaptive filter is only updated during noise-only-periods.

Algorithm	Complexity	MIPS
GSC-SPA	$(3M-1)\mathrm{FFT}+14M-12$	2.02
MWF (Algo1)	$(3N+5)\mathrm{FFT}+28N+6$	3.10$^{(a)}$, 4.13$^{(b)}$
MWF (Algo3)	$(3N+2)\mathrm{FFT}+8N^2+14N+3$	2.54$^{(a)}$, 3.98$^{(b)}$
	Memory	**kWords**
GSC-SPA	$4(M-1)L+6L$	0.45
MWF (Algo1)	$2NL_y+6LN+7L$	40.61$^{(a)}$, 60.80$^{(b)}$
MWF (Algo3)	$4LN^2+6LN+7L$	1.12$^{(a)}$, 1.95$^{(b)}$

Table 1: Computational complexity and memory usage for $M=3$, $L=32$, $f_s = 16$ kHz, $L_y = 10000$, (a) $N = M-1$, (b) $N = M$

(assuming the radix-2 FFT algorithm). From this table we can draw the following conclusions:

- The *computational complexity* of the SDW-MWF (Algorithm 1) with filter $\mathbf{w}_0$ is about twice the complexity of the GSC-SPA (and even less without $\mathbf{w}_0$). The approximation in the SDW-MWF (Algorithm 3) further reduces the complexity. However, this only remains true for a small number of input channels, since the approximation introduces a quadratic term $\mathcal{O}(N^2)$.
- Due to the storage of the speech+noise-buffer, the *memory usage* of the SDW-MWF (Algorithm 1) is quite high in comparison with the GSC-SPA (depending on the size of the data buffer L_y of course). By using the approximation in the SDW-MWF (Algorithm 3), the memory usage can be drastically reduced. Note however that also for the memory usage a quadratic term $\mathcal{O}(N^2)$ is introduced.

5. EXPERIMENTAL RESULTS

In this paragraph it is shown that practically no performance difference exists between implementing the SDW-MWF using Algorithm 1 or 3, such that the SDW-MWF using the proposed implementation preserves its robustness benefit over the GSC (and the QIC-GSC).

5.1 Set-up and performance measures

A 3-microphone BTE has been mounted on a dummy head in an office room. The desired source is positioned in front of the head (at 0°) and consists of English sentences. The noise scenario consists of three multi-talker babble noise sources, positioned at 75°, 180° and 240°. The desired signal and the total noise signal both have a level of 70 dB SPL at the centre of the head. For evaluation purposes, the speech and the noise signal have been recorded separately. In the experiments, the microphones have been calibrated in an anechoic room with the BTE mounted on the head. A delay-and-sum beamformer is used as fixed beamformer $\mathbf{A}(z)$. The blocking matrix $\mathbf{B}(z)$ pairwise subtracts the time-aligned calibrated microphone signals. The filter length $L = 32$, the step size $\rho' = 0.8$, $\gamma = 0.95$ and $\lambda = 0.999$.

To assess the performance, the intelligibility weighted signal-to-noise ratio improvement $\Delta\mathrm{SNR}_{\mathrm{intellig}}$ is used, defined as

$$\Delta\mathrm{SNR}_{\mathrm{intellig}} = \sum_i I_i(\mathrm{SNR}_{i,\mathrm{out}} - \mathrm{SNR}_{i,\mathrm{in}}), \tag{13}$$

where I_i expresses the importance for intelligibility of the i-th one-third octave band with centre frequency f_i^c [12], and where $\mathrm{SNR}_{i,\mathrm{out}}$ and $\mathrm{SNR}_{i,in}$ are respectively the output and the input SNR (in dB) in this band. Similarly, we define an intelligibility weighted spectral distortion measure, called $\mathrm{SD}_{\mathrm{intellig}}$, of the desired signal as

$$\mathrm{SD}_{\mathrm{intellig}} = \sum_i I_i \mathrm{SD}_i \tag{14}$$

with SD_i the average spectral distortion (dB) in the i-th one-third band, calculated as

$$\mathrm{SD}_i = \frac{1}{\left(2^{1/6} - 2^{-1/6}\right) f_i^c}\int_{2^{-1/6}f_i^c}^{2^{1/6}f_i^c} |10\log_{10} G_x(f)|\, df, \tag{15}$$

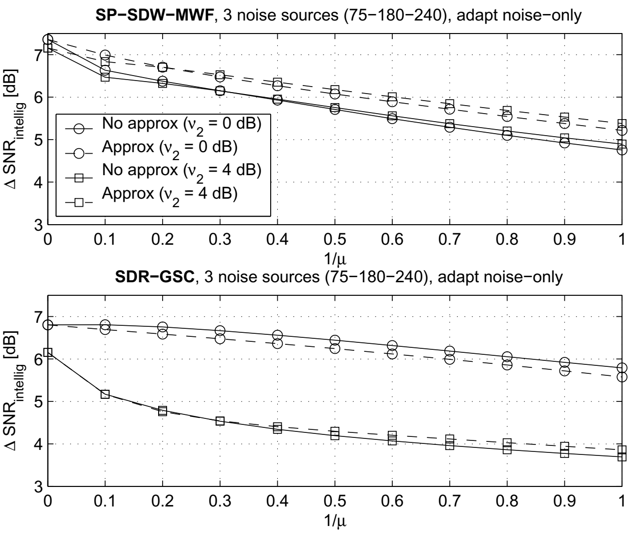

Figure 2: SNR improvement of FD SP-SDW-MWF (with and without approximation) in a multiple noise source scenario

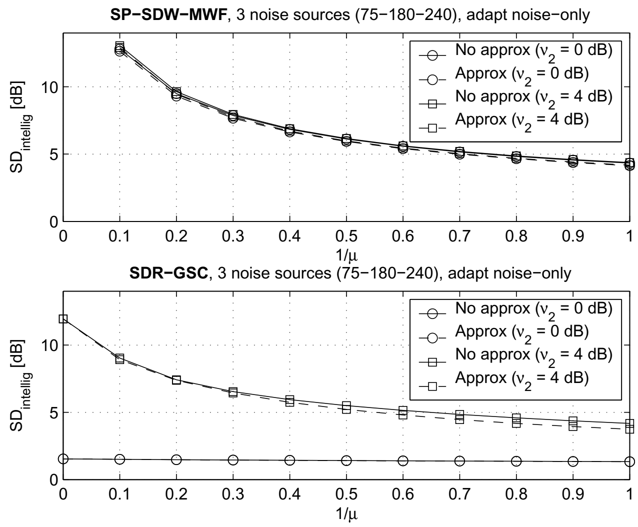

Figure 3: Speech distortion of FD SP-SDW-MWF (with and without approximation in a multiple noise source scenario

with $G_x(f)$ the power transfer function of speech from the input to the output of the noise reduction algorithm. To exclude the effect of the spatial pre-processor, the performance measures are calculated w.r.t. the output of the fixed beamformer, i.e. the speech reference.

5.2 Experimental results

Figures 2 and 3 depict the SNR improvement and the speech distortion of the SP-SDW-MWF (with $\mathbf{w}_0$) and the SDR-GSC (without $\mathbf{w}_0$) as a function of the trade-off parameter $1/\mu$, for Algorithm 1 (no approx) and Algorithm 3 (approx). These figures also depict the effect of a gain mismatch $\nu_2 = 4\,\text{dB}$ at the second microphone. From these figures it can be observed that approximating the regularisation term results in a small performance difference (smaller than 0.5 dB). For some scenarios the performance is even better for Algorithm 3 than for Algorithm 1, probably since in Algorithm 1 it is assumed that the filter $\mathbf{w}[k]$ varies slowly in time.

Hence, also when implementing the SDW-MWF using Algorithm 3, it still preserves its robustness benefit over the GSC (and the QIC-GSC). E.g. it can be observed that the GSC (i.e. SDR-GSC with $1/\mu = 0$) will result in a large speech distortion (and a smaller SNR improvement) when microphone mismatch occurs. Both the SDR-GSC and the SDW-MWF add robustness to the GSC, i.e. distortion increases for increasing $1/\mu$. The performance of the SDW-MWF is even hardly effected by microphone mismatch.

6. CONCLUSION

In this paper we have shown that the memory usage (and the computational complexity) of the SDW-MWF can be reduced drastically by approximating the regularisation term in the frequency-domain, i.e. by computing the regularisation term using (diagonal) FD correlation matrices instead of TD data buffers. It has been shown that approximating the regularisation term only results in a small performance difference, such that the robustness benefit of the SDW-MWF is preserved at a smaller computational cost, which is comparable to the NLMS-based implementation for QIC-GSC.

REFERENCES

[1] A. Spriet, M. Moonen, and J. Wouters, "Stochastic gradient implementation of spatially pre-processed multi-channel Wiener filtering for noise reduction in hearing aids," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.

[2] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction in hearing aids," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Japan, Sept. 2003, pp. 147–150.

[3] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Proc.*, vol. 50, no. 9, pp. 2230–2244, Sept. 2002.

[4] G. Rombouts and M. Moonen, "QRD-based unconstrained optimal filtering for acoustic noise reduction," *Signal Processing*, vol. 83, no. 9, pp. 1889–1904, Sept. 2003.

[5] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust Adaptive Beamforming," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 10, pp. 1365–1376, Oct. 1987.

[6] J. E. Greenberg and P. M. Zurek, "Evaluation of an adaptive beamforming method for hearing aids," *Journal of Acoust. Soc. of America*, vol. 91, no. 3, pp. 1662–1676, Mar. 1992.

[7] S. Nordholm, I. Claesson, and B. Bengtsson, "Adaptive Array Noise Suppression of Handsfree Speaker Input in Cars," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 514–518, Nov. 1993.

[8] D. A. Florêncio and H. S. Malvar, "Multichannel filtering for optimum noise reduction in microphone arrays," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City UT, USA, May 2001, pp. 197–200.

[9] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Proc. Magazine*, pp. 15–37, Jan. 1992.

[10] J. Benesty and D. R. Morgan, "Frequency-domain adaptive filtering revisited, generalization to the multi-channel case, and application to acoustic echo cancellation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, May 2000, pp. 789–792.

[11] R. Aichner, W. Herbordt, H. Buchner, and W. Kellermann, "Least-squares error beamforming using minimum statistics and multichannel frequency-domain adaptive filtering," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Japan, Sept. 2003, pp. 223–226.

[12] Acoustical Society of America, "ANSI S3.5-1997 American National Standard Methods for Calculation of the Speech Intelligibility Index," June 1997.